'Dynamics of neuronal interactions' cannot be explained by 'neuronal transients'

E. VAADIA[1]*, A. AERTSEN[2] AND I. NELKEN[3]

[1] *Department of Physiology, Hadassah School of Medicine & The Center for Neural Computation, The Hebrew University, Jerusalem, 91120 Israel*
[2] *The Center for Research of Higher Brain Functions, Department of Neurobiology, The Weizmann Institute of Science, Rehovot, 76100 Israel*
[3] *Department of Physiology, Hadassah School of Medicine, The Hebrew University, Jerusalem, 91120 Israel*

SUMMARY

In a recent paper, Vaadia *et al.* demonstrated that patterns of firing correlation between single neurons in the cortex of behaving monkeys can be modified within a fraction of a second. These changes occur in relation to sensory stimuli and behavioral events, and even without modulations of the neurons' firing rates. These findings call for a revision of prevailing models of neural coding that solely rely on single neuron firing rates. In a defense of these models, Friston put forward an alternative explanation, proposing that the observed correlation dynamics emerge solely from co-modulations of the firing rates of each of the neurons, while the strength of their interaction remains constant. To test this possibility we re-examined the data, adopting Friston's 'neuronal transients' model, and the associated equations and procedures. We found that, to explain the dynamic correlation between a pair of neurons, the alternative interpretation requires that each neuron's response to a single stimulus is composed of a relatively large number of independent components, which co-vary with their counterparts in the companion neuron. This large number of components and their shapes lead us to conclude that, although in principle possible, the neuronal transients model: (i) does not provide a simpler explanation of the experimental results; and (ii) cannot explain these results without itself deviating significantly from most rate code models.

In his paper 'Neuronal transients' Friston (1995) presents an alternative explanation of the results of the joint peri-stimulus time (JPST) analysis for the spike trains of pairs of simultaneously recorded frontal cortex neurons by Vaadia *et al.* (1995). One of the prime goals of his explanation is to 'reinstate' prevailing models of neuronal coding based on firing rates. To test if the neuronal transient model provides an adequate explanation for our experimental findings, we re-analysed the data we presented by using the analysis and equations (1), (2) and (4) in Friston's paper (Friston 1995). First, we estimated the 'non-specific correlation'. Then, by using singular value decompostition (SVD) analysis (which is a generalization of principal components analysis to non-symmetric and not-square matrices; for more information see Press *et al.* 1988), we computed the hypothetical 'neuronal transients' that best explain the data. Finally, we computed the correlation matrix, generated by the weighted contributions of the 'neuronal transients' and the 'non-specific correlation' (see legend of figure 1 for details of this inverse approach). We found that the neuronal transients model does not provide a simple explanation of our results.

* Correspondence address: Department of Physiology, Hadassah School of Medicine, P.O.B. 12272, Jerusalem, Israel 91120.

The results of the analysis for the spike trains of one neuronal pair from figure 2*a* of Vaadia *et al.* (1995) are compared to the results of JPST analysis in figures 1–3. We first examined the 'non-specific' (constant) correlation. By its stationary nature, it cannot account for dynamic modulation of correlation along the diagonal. Rather, it changes the excess correlation along the diagonal by a constant amount, without affecting its time course. This is illustrated in figure 1 for the example discussed here. The solid curve replicates the coincidence-time histogram from our JPST-matrix (the green histogram in Vaadia *et al.* (1995), see figure 2*a*). The correlation gradually builds up, starting 1000 ms before the stimulus, and reaches a peak about 400 ms after it. After subtraction of the estimated 'non-specific' correlation (dotted line), the correlation dynamics maintain a virtually identical time course (dashed line). Thus, Friston's proposition that non-specific correlation would account for 'marked correlation before the stimulus arrives' is not substantiated. We note in passing that event-related 'anticipatory' activity (be it rate or correlation) preceding stimuli or other behavioral events is frequently encountered in the frontal cortex of behaving monkeys (Tanji & Evarts 1976; Vaadia *et al.* 1988).

To evaluate the combined contributions of non-specific correlation and co-varying 'neuronal transi-

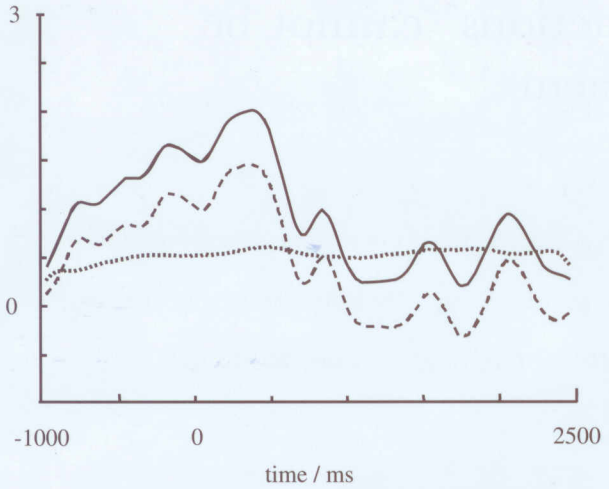

Figure 1. Dynamic modulation of correlated firing. The solid line replicates the diagonal (green) coincidence-time histogram in Vaadia *et al.* (1995) figure 2*a*. The dotted line is the estimated 'non-specific correlation' for the same data. The difference between these two graphs is shown in the dashed line. The small fluctuations of the 'non-specific' correlations are the result of fluctuations in the estimated standard deviation (the non-specific covariance is constant, but we divided it by the standard deviation to obtain the non-specific correlation as follows. We re-analysed the data in Vaadia *et al.*(1995), figure 2*a*, using SVD-analysis and the model equations (1), (2) and (4) in Friston (1995). Since, unlike Friston, because we based our analysis on physiological data, the values of the 'non-specific activities' e_i and e_j were not available. Therefore the non-specific covariance (γ_{ij}) was estimated by computing the covariance between the two spike trains over the time window between 1500–2500 ms after the 'ready' signal (cf. figure 3*a*). During this interval there is no discernible modulation in the JPST correlation matrix. This estimate, extended along the entire diagonal, was subtracted from the covariance matrix (the raw JPST minus the PST-predictor (Aertsen *et al.* 1989)), and the resulting difference matrix was subjected to SVD-analysis. The first pair of singular vectors (see figure 2, top traces) were taken as estimators of the hypothetical 'rate transients' (τ_i and τ_j). To produce the prediction of the SVD-analysis for our normalized correlation matrix, we first evaluated the contribution of the first term in equation (2) of Friston (1995) (the product of the first singular vectors, multiplied by the corresponding singular value). We then added back the non-specific covariance (γ_{ij}). Finally, we divided the resulting covariance matrix by the standard deviation matrix (Aertsen *et al.* 1989) and obtained the predicted correlation matrix (figure 3*b*). We note that there are practical difficulties with estimating the 'nonspecific correlation' from spike train data. Experimental observations show different amounts of correlation under different conditions and at different times, even before or in the absence of a stimulus. This is also the case in the present example. Had we taken, for example, the 1000 ms preceding the ready signal in figure 3*a* to estimate γ_{ij}, we would have arrived at a higher non-specific correlation. This, however, would mainly shift the dashed line downwards, and hardly affect its time course.)

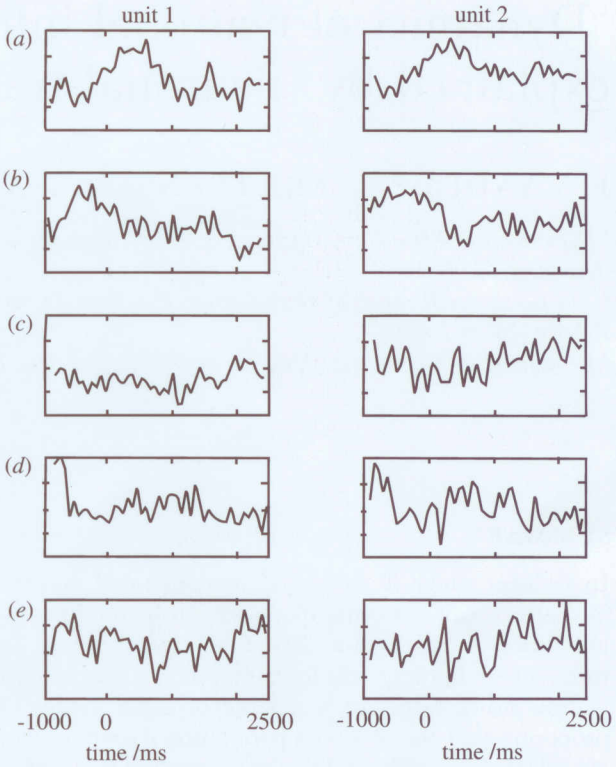

Figure 2. The first five pairs of singular vectors computed by SVD-analysis of the spike train data presented in figure 2*a* of Vaadia *et al.* (1995). The corresponding singular values decay slowly (λ_1–λ_5: 256, 227, 200, 169, 148). Accordingly, there is only a slow increase in the variance accounted for by the first few singular vectors (the fraction of the variance contributed by each of the first five vectors are: 17%, 14%, 10%, 8%, 6%).

ents', we constructed the correlation matrix in figure 3*b*, out of the first pair of singular vectors (shown in figure 2, top traces). This matrix should be compared with the normalized JPST matrix in figure 3*a*. The 'neuronal transient' model predicts a constant diagonal (the 'non-specific correlation') with a wide square 'hot spot' around time 500 ms (see figure 3*b*), corresponding to the peak of the JPST matrix (figure 3*a*). By contrast, the main feature of the JPST matrix is the gradual, elongated build-up of correlation along the diagonal, described above (cf. solid curve in figure 1). This elongated feature is conspicuously absent in figure 3*b*. In fact, such elongated features can never be explained by a single transient. Overall, figure 3*b* accounts for only 17% of the variance in figure 3*a*. Full reconstruction of the normalized JPST would require additional transients, contributing multiple hot spots along the diagonal (strung like beads on a chain). Indeed, additional singular vectors gradually improve the performance of the neuronal transient model. Visual inspection suggests that between 5–10 singular vectors are required to recreate the diagonal buildup: five vectors (shown in figure 2) explain 55% of the variance in figure 3*a*, ten vectors are needed to account for 74%. Clearly, such large numbers of transients are hard to reconcile with the conventional notion of rate coding, which assumes that the firing rate, averaged over some appropriate time window, carries the information. There have been indeed proposals to broaden this definition to incorporate also the time course (waveform) of neuronal responses. In these cases, principal components analysis revealed that the meaningful information was contained in the first three

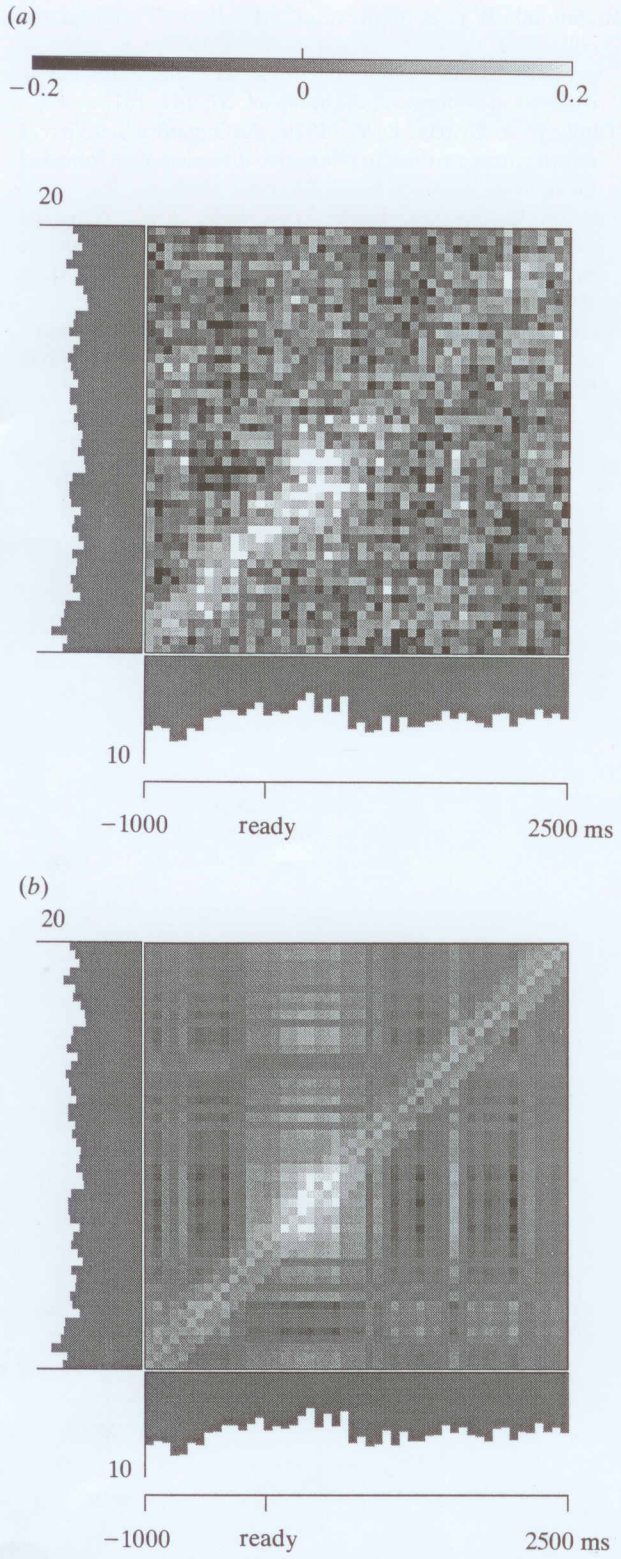

Figure 3. (*a*) Normalized JPST correlation matrix from figure 2*a* of Vaadia *et al.* (1995). (*b*) Estimation of the correlation matrix in (*a*), based on the combined contributions of non-specific correlation and the first pair of singular vectors weighted by the corresponding singular value (constructed as described in the legend of figure 1).

components (Richmond & Optican 1987), whereas in the present case, even ten components are not sufficient. Moreover, examining the nondescript waveforms of the higher order neuronal transients (figure 2*c–e*), we find it hard to associate them with any consistent modulation of the firing rates, let alone to a physiological process that would be linked with them.

Thus, our analysis indicates that the neuronal transients model can explain the experimental data only if four conditions are met: (i) the responses of the two neurons i and j to a specific stimulus are characterized by 5–10 different 'transients' (firing rate modulations τ_i^k and τ_j^k, $k = 1$–10); (ii) each stimulus presentation evokes a mixture of these transients with random weight factors α_i^k and α_j^k; (iii) the weight factors of the two neurons are uncorrelated for different k's; but, (iv) they are correlated for the same k (the singular value λ^k is a measure of this correlation). We conclude that Friston's alternative interpretation is, in principle, possible but is not 'a simpler explanation' (Friston 1995) of our results.

Under these conditions, and particularly in view of the need to invoke relatively large numbers of transients, the neuronal transients model deviates significantly from most rate code models. Also, preliminary SVD-analysis of pair correlations among a simultaneously recorded triplet of neurons suggests that the dominant singular vectors of a given neuron may be different, depending on the selected 'partner' neuron. This would imply that the 'transients' are not only a property of the single neuron, but also of the selected pair. Evidently, this is not what could be expected from a rate code model.

Nevertheless, we do believe that Friston makes two important points, that provide valuable suggestions for future studies.

1. SVD-analysis is certainly a helpful tool for the analysis of neuronal activity. It could be used to evaluate typical temporal patterns of firing rate modulations, and to characterize and quantify the variability of these modulations across repeated presentations of the same stimulus. A combination of SVD-analysis with measurements of spike correlation can be used to assess the relative contributions of these modulations to the neural code.

2. The incorporation of 'nonspecific correlation' provides an attractive alternative null hypothesis for the study of neuronal interactions. Rather than to test against the classical null hypothesis of zero correlation (independent firing) as the current JPST-analysis does, it might indeed be more appropriate (stronger) to test for the presence of correlation dynamics against the null hypothesis of constant correlation, provided a good experimental estimate of this constant correlation can be obtained.

In spite of this, however, the prime objective of Friston's proposal is not met. The potential attraction of his model would be that it might explain our experimental results without the need to go beyond the widely accepted firing rate models. As we have demonstrated, however, this appears to be impossible, without deviating from the conventional rate model. We maintain that a more attractive alternative is, in fact, to venture beyond single neuron rate coding and to incorporate the correlation of spike firing as a promising new mechanism to dynamically organize neurons into functional groups.

We thank M. Abeles, H. Bergman, I. Haalman, Y. Prut and H. Slovin for their contributions. The study was supported in part by grants from the Israeli Academy of Sciences and Humanities, the American-Israel Bi-national Science Foundation (BSF), and the 'Human Frontier Science Program (HFSP)'.

REFERENCES

Aertsen, A. M., Gerstein, G. L., Habib, M. K. & Palm, G. 1989 Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. *J Neurophysiol.* **61**, 900–917.

Friston, K. J. 1995 Neuronal transients. *Proc. R. Soc. Lond.* B **262**, 000–000.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. 1988 Numerical references in C. In *The art of scientific computing*, pp. 60–72. Cambridge University Press.

Richmond, B. J. & Optican, L. M. 1987 Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol.* **57**, 147–161.

Tanji, J. & Evarts, E. V. 1976 Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J. Neurophysiol.* **39**, 1062–1068.

Vaadia, E., Kurata, K. & Wise, S. P. 1988 Neuronal activity preceding directional and nondirectional cues in the premotor cortex of rhesus monkeys. *Somatosensory Motor Res.* **6**, 207–230.

Vaadia, E., Haalman, I., Abeles, M. *et al.* 1995 Dynamics of neuronal interactions in the monkey cortex in relation to behavioral events. *Nature, Lond.* **373**, 515–518.

Received 2 June 1995; accepted 20 June 1995